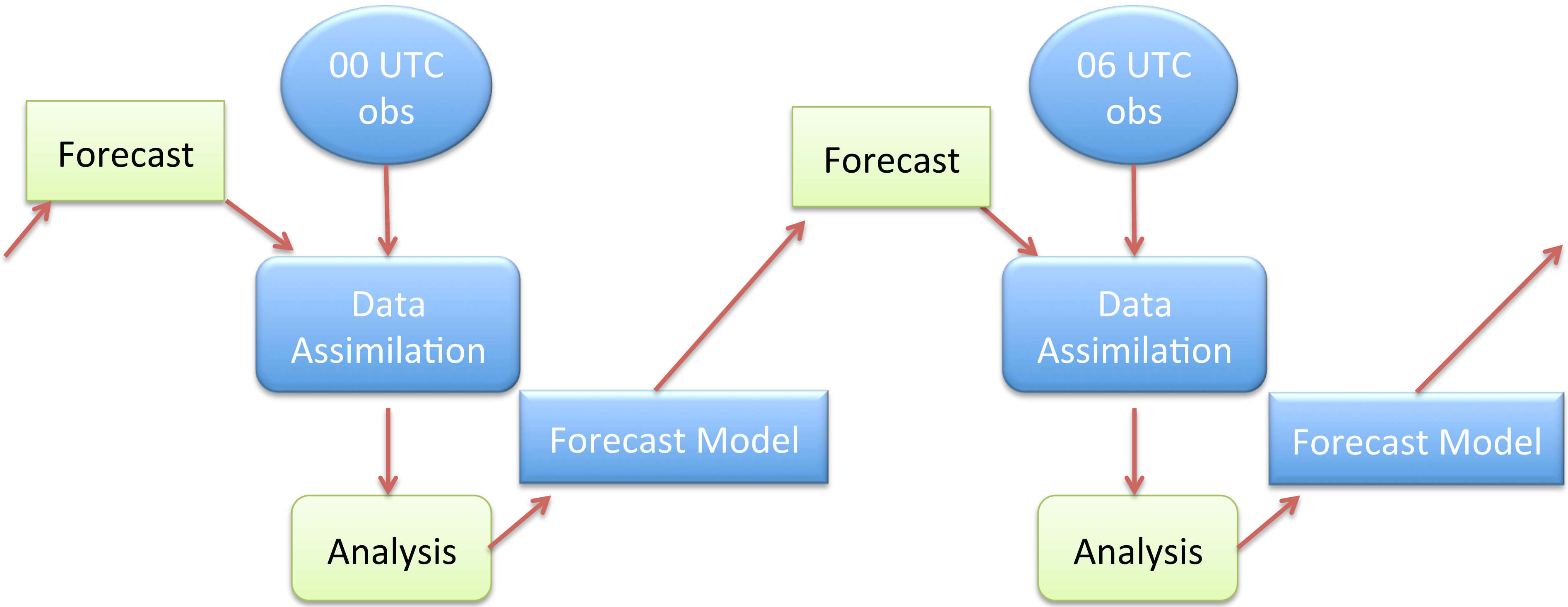
Using Ensemble Forecasts to Improve Data Assimilation for Weather Prediction

Jeff Whitaker jeffrey.s.whitaker@noaa.gov

Tom Hamill, Phil Pegion (ESRL/PSD)
Daryl Kleist, Dave Parrish, John Derber (NWS/NCEP)

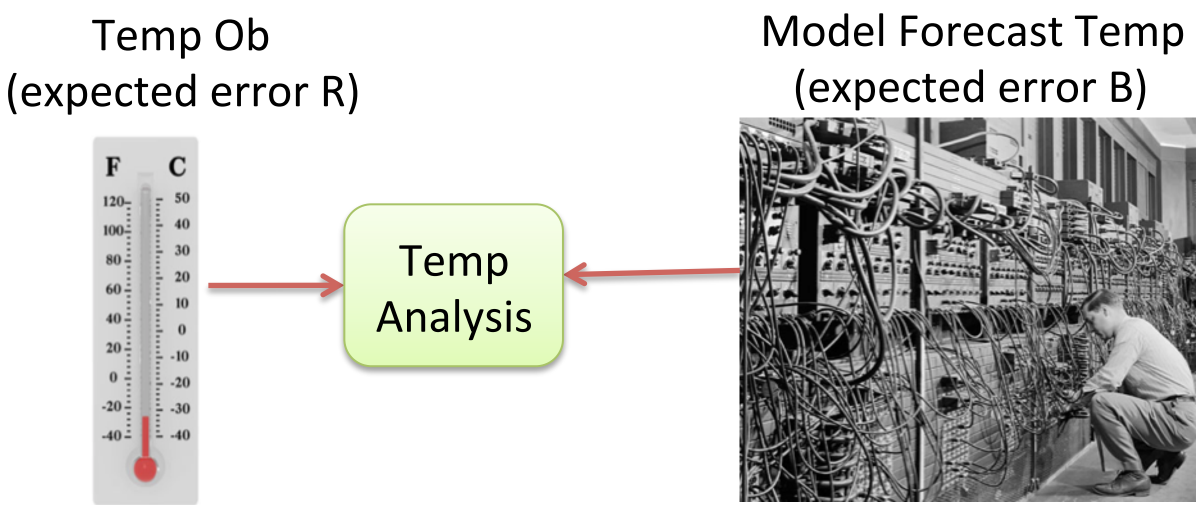
The Numerical Weather Prediction Process



- Analyses and forecasts become more accurate when:
  - Observations, forecast model and/or data assimilation components improve.
  - Forecast model carries information from past observations.

What is Data Assimilation?

The processing of combining prior knowledge of the state of the atmosphere (a previous model forecast) with new observations.



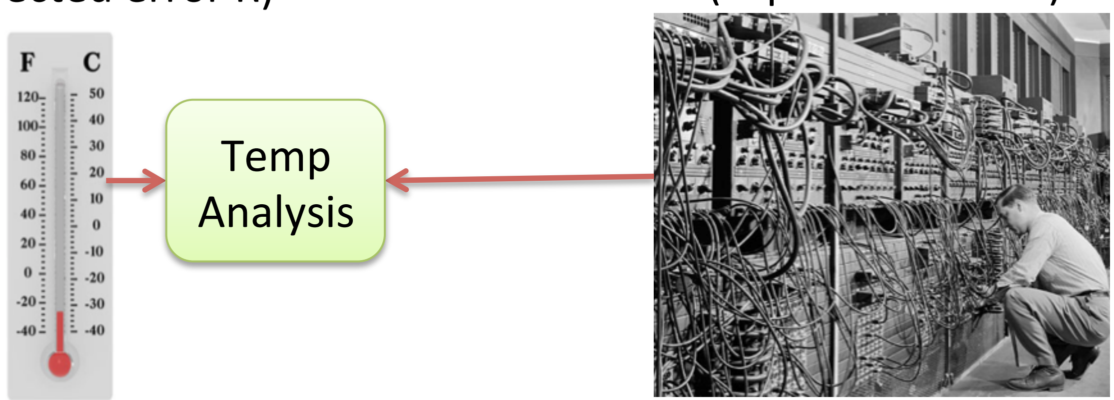
If we have equal confidence in the prior forecast and a new observation (R = B), analysis is half-way between (equal weight given to each).

What is Data Assimilation?

The processing of combining prior knowledge of the state of the atmosphere (a previous model forecast) with new observations.

Temp Ob
(expected error R)

Model Forecast Temp
(expected error B)



If we have less confidence in the prior forecast (B > R), analysis is closer to the observation. *The job of DA is to compute the weights that optimally blend the observations and the model forecast.*

Data assimilation terminology

- **y** : Observation vector (weather balloons, satellite radiances, etc.)
- **x** : the state of the atmosphere as represented by the model
- $\mathbf{x}^b$: Background state vector ("prior")
- $\mathbf{x}^a$: Analysis state vector ("posterior")
- **H** : (hopefully linear) operator to convert model state → observation location & type
- **R** : Observation - error covariance matrix
- $\mathbf{P}^b$: Background - error covariance matrix
- $\mathbf{P}^a$: Analysis - error covariance matrix

Bayesian Data Assimilation (assuming Gaussian PDFs)

Assume errors in first-guess forecast are Gaussian, with mean $\mathbf{x}_b$ and covariance $\mathbf{P}^b$.

$$p(\mathbf{x}) \propto \exp\left(-(\mathbf{x}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{P}^{\mathrm{b}^{-1}}(\mathbf{x}-\mathbf{x}_\mathrm{b})\right)$$

Conditional prob of obs ($\mathbf{y}$) given state ($\mathbf{x}$), which has mean $\mathbf{Hx}$ and covariance $\mathbf{R}$.

$$p(\mathbf{y}|\mathbf{x}) \propto \exp\left(-(\mathbf{y}-\mathbf{Hx})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Hx})\right)$$

So, substituting into Bayes Rule yields the conditional prob. of state given past observations…

$$p(\mathbf{x}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{x})p(\mathbf{x})$$

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-(\mathbf{x}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{P}^{\mathrm{b}^{-1}}(\mathbf{x}-\mathbf{x}_\mathrm{b}) - (\mathbf{y}-\mathbf{Hx})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{Hx})\right)$$

From Bayes theorem to 4DVar and the (Ensemble) Kalman Filter

$$p(\mathbf{x}|\mathbf{y}) \propto \exp\left(-(\mathbf{x}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{P}^{\mathrm{b}^{-1}}(\mathbf{x}-\mathbf{x}_\mathrm{b}) - (\mathbf{y}-\mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})\right)$$

Variational methods maximize the posterior PDF to find the state trajectory **x** that best fits the obs **y** in a least-squares sense. In practice, this is done by minimizing a cost function, which is what's inside the *exp*:

$$J(\mathbf{x}) \propto (\mathbf{x}-\mathbf{x}_\mathrm{b})^\mathrm{T}\mathbf{P}^{\mathrm{b}^{-1}}(\mathbf{x}-\mathbf{x}_\mathrm{b}) + (\mathbf{y}-\mathbf{H}\mathbf{x})^\mathrm{T}\mathbf{R}^{-1}(\mathbf{y}-\mathbf{H}\mathbf{x})$$

The minimum can be found analytically if **H** is linear (see Lorenc 1986 *QJRMS* for the algebra). This gives the equations for the Kalman Filter

$$\boxed{\begin{gathered}\mathbf{x}_\mathrm{a} = \mathbf{x}_\mathrm{b} + \mathbf{K}\left(\mathbf{y}-\mathbf{H}\mathbf{x}_\mathrm{b}\right),\ \mathbf{P}^\mathrm{a} = \left(\mathbf{I}-\mathbf{K}\mathbf{H}\right)\mathbf{P}^\mathrm{b} \\ \mathbf{K} = \mathbf{P}^\mathrm{b}\mathbf{H}^\mathrm{T}\left(\mathbf{H}\mathbf{P}^\mathrm{b}\mathbf{H}^\mathrm{T}+\mathbf{R}\right)^{-1}\end{gathered}}$$

- *Matrix* $\mathbf{P}^\mathrm{b}$ *is too big for any computer, covariance update step impractical.*

- *Instead, represent PDFs of* **x** *and* **y** *by an ensemble, compute sample estimate of* $\mathbf{P}^\mathrm{b}$*. Evolve the sample, not the full covariance.* ***EnKF*** *gives same result as full KF if ensemble size becomes infinite.*

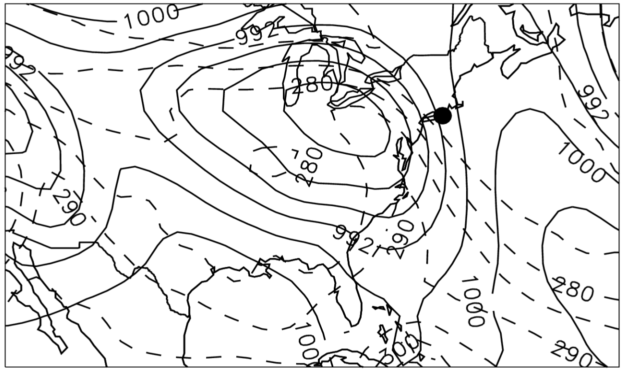
$\mathbf{P}^b$ in the Variational Solution

- "Assimilation window": time interval that encompasses observations to be assimilated (6 hours to a day).
- "3DVar": Assume $\mathbf{P}^b$ is does not vary in assimilation window (climatological estimate).
- "4DVar": $\mathbf{P}^b$ specified at beginning of assim. window, evolved with linearized model through the window. Accuracy of covariance estimate depends on:
  - Accuracy of linear model (typically low-resolution with very simple physics).
  - Accuracy of initial condition (initial $\mathbf{P}^b$ typically set to 3DVar version at start of window), length of window.

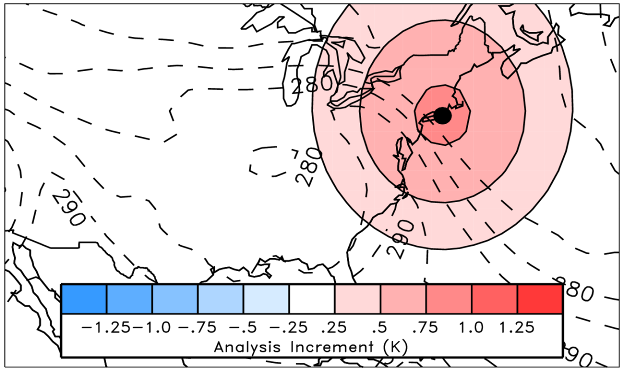
Why does P^b matter?
Example 1: Front

Temperature observation near a warm front

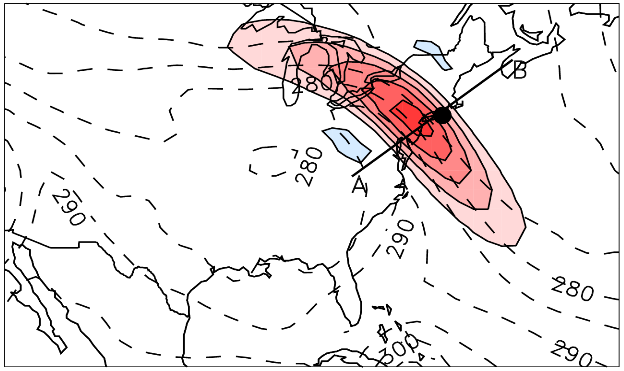
1000 hPa temperature (K) and surface pressure (hPa)



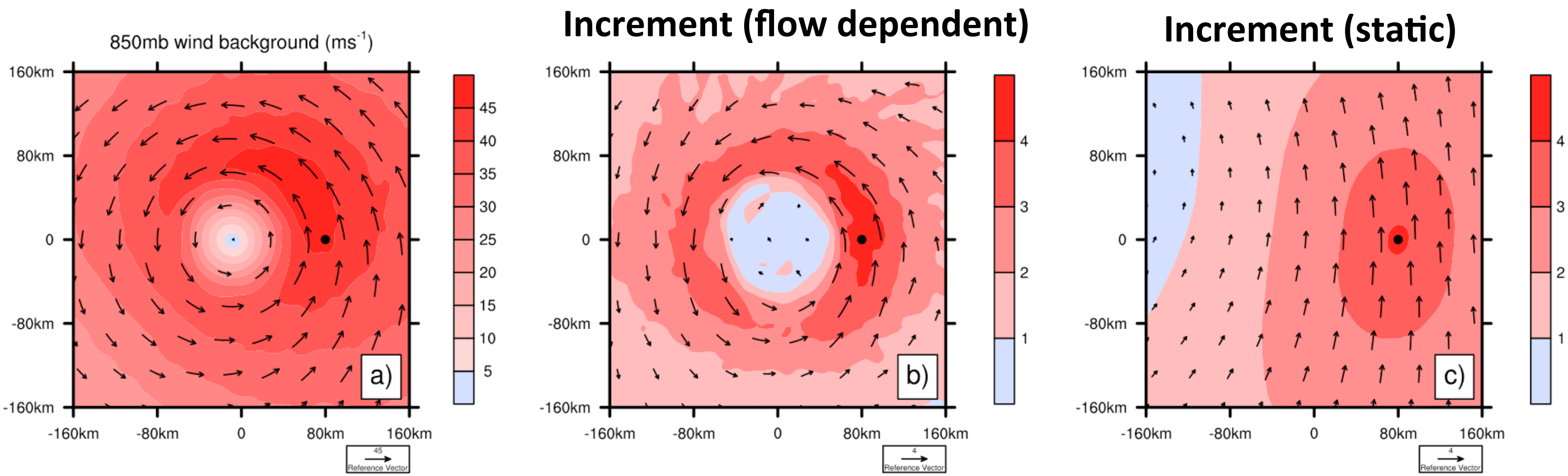
Increment (static)



Increment (flow dependent)



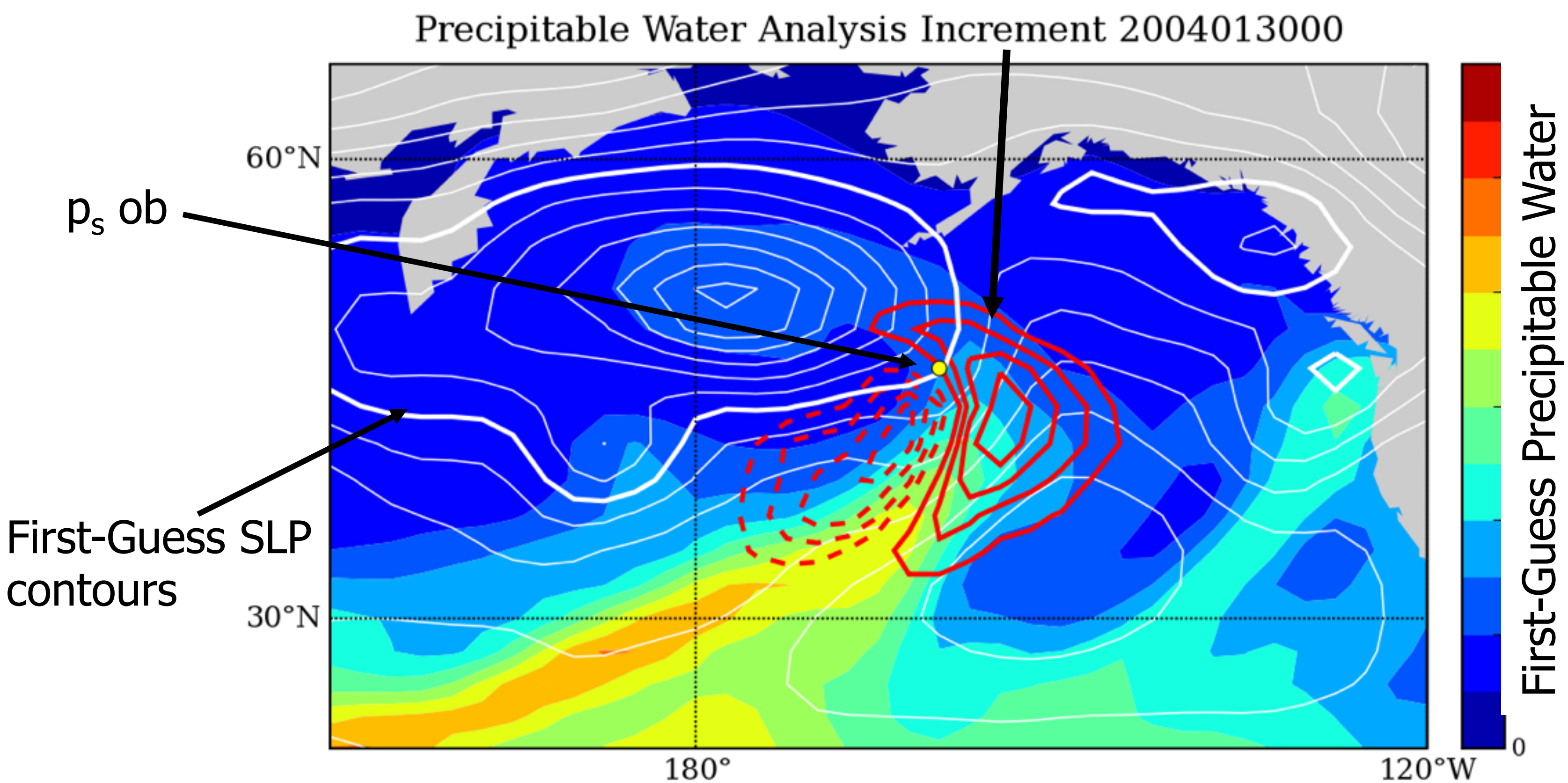
Why does P^b matter?
Example 2: Hurricane



Flow dependant covariances in a hurricane can produce axisymmetric increments for a single ob.

figure c/o Xuguang Wang, University of Oklahoma

Why does P^b matter?
Example 3: unobserved variables



Surface pressure observation can improve analysis of integrated water vapor (through flow-dependent cross-variable relationships). If climo weighting were used (3DVar) there would be essentially no vapor increment.

Computational shortcuts in EnKF:
(1) Simplifying Kalman gain calculation

$$\mathbf{K} = \mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} \left(H\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} + \mathbf{R} \right)^{-1}$$

$$\textit{define} \quad \overline{H\mathbf{x}^{\mathrm{b}}} = \frac{1}{m} \sum_{i=1}^{m} H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}}$$

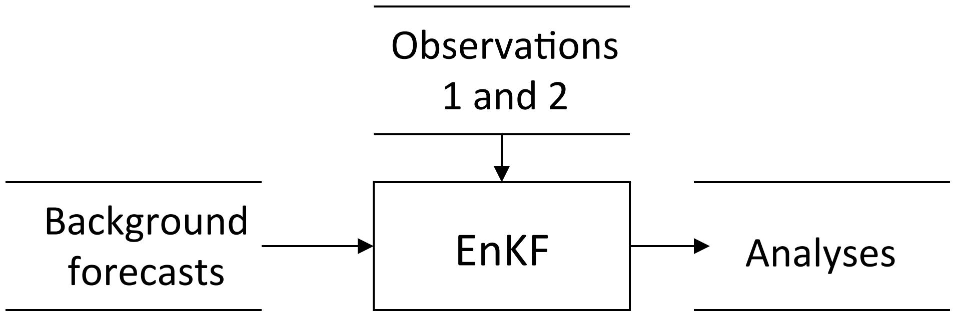
$$\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{\mathbf{x}^{\mathrm{b}}} \right) \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right)^{\mathrm{T}}$$

$$H\mathbf{P}^{\mathrm{b}} H^{\mathrm{T}} = \frac{1}{m-1} \sum_{i=1}^{m} \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right) \left(H\mathbf{x}_{\mathrm{i}}^{\mathrm{b}} - \overline{H\mathbf{x}^{\mathrm{b}}} \right)^{\mathrm{T}}$$

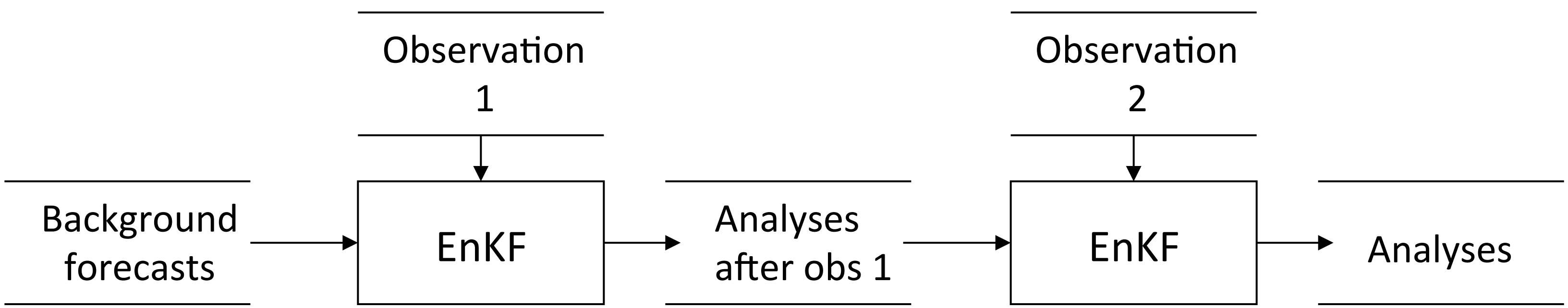
The key here is that the huge matrix $\mathbf{P}^{\mathrm{b}}$ is never explicitly formed

Computational shortcuts in EnKF:
(2) serial processing of observations (requires observation error covariance **R** to be diagonal)

Method 1



Method 2



The serial EnKF – a recipe

Given a single ob y^o with expected error variance R, an ensemble of model forecasts $\mathbf{x}^b$ (model priors), and an ensemble of predicted observations $\mathbf{y}^b = \mathbf{H}\mathbf{x}^b$ (observation priors):

Step 1: Update observation priors.

(1a) $\overline{\mathbf{y}}_a = (1-K)\overline{\mathbf{y}}_b + Ky^o$ *update for ob prior means*

(1b) $\mathbf{y}_a^{'} = \sqrt{(1-K)}\mathbf{y}_b^{'}$ *rescaling of ob prior perturbations*

where the scalar K = var($\mathbf{y}^b$)/(var($\mathbf{y}^b$)+ R), overbar denotes means, prime denotes perturbations, superscript *b* denotes prior, *a* denotes analysis.

Linear interpolation between observation and observation prior mean *with weight K (0<=K<=1), rescaling of observation prior ensemble so posterior variance is consistent with Kalman filter, i.e.* var($\mathbf{y}^a$)=(1-K) var($\mathbf{y}^b$)= var($\mathbf{y}^b$)R/(var($\mathbf{y}^b$)+R).

when var($\mathbf{y}^b$) << R, all weight given to prior.

when var($\mathbf{y}^b$) >> R, all weight given to observation.

The serial EnKF – a recipe (2)

Step 2: Update model priors.

Let $\Delta\mathbf{x} = \mathbf{x}^a - \mathbf{x}^b$ be analysis increment for model priors, $\Delta\mathbf{y} = \mathbf{y}^a - \mathbf{y}^b$ is analysis increment for observation priors.

(2) $\Delta\mathbf{x} = \mathbf{G}\Delta\mathbf{y}$ *computation of increments to model prior*

where $\mathbf{G} = \mathrm{cov}(\mathbf{x}^b, \mathbf{y}^{bT})/\mathrm{var}(\mathbf{y}^b)$

Linear regression of model priors on observation priors.

Only changes model priors when $\mathbf{x}^b$ and $\mathbf{y}^b$ are correlated within the ensemble.

If there is more than one ob to be assimilated, the observation priors for other (not yet assimilated) obs ($\mathbf{Y}^b$) should be also be updated using (2) with $\Delta\mathbf{x}$ replaced by $\Delta\mathbf{Y}$. Next iteration, replace $\mathbf{y}^b$ with next column of $\mathbf{Y}^b$, removing that column from $\mathbf{Y}^b$. After each iteration the model priors and observation priors are set to the latest analysis values ($\mathbf{x}^a$ replaces $\mathbf{x}^b$, $\mathbf{Y}^a$ replaces $\mathbf{Y}^b$). Continue iterating until $\mathbf{Y}^b$ is empty.

Factors limiting EnKF performance

1) Treatment of model error

Must account for the background error covariance associated with "model error" (any difference between simulated and true environment). Methods used so far:

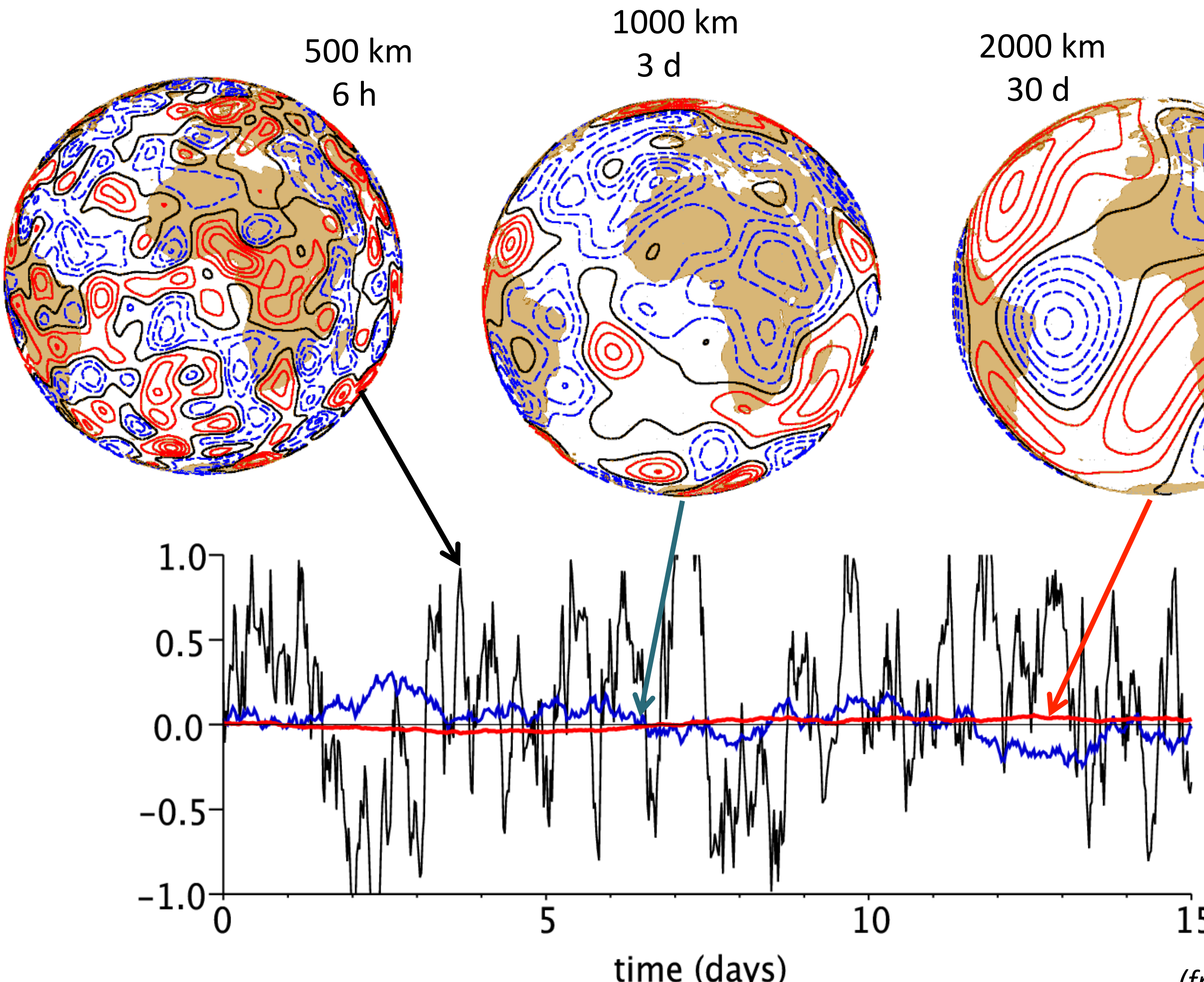
1) multiplicative inflation (mult. ens perts by a factor > 1).
2) additive inflation (random perts added to each member – e.g. differences between 24 and 48-h forecasts valid at the same time).
3) model-based schemes (e.g. stochastic kinetic energy backscatter for representing unresolved processes, multi-model/multi-parameterization).

Opnl NCEP system uses a combination of 1) and 2).

Can we replace the additive inflation by adding stochastic physics to the model?

- Schemes tested:
  - SPPT (stochastically perturbed physics tendencies)
  - SKEB (stochastic KE backscatter)
  - VC (vorticity confinement, *deterministic and/or stochastic)*
  - SHUM (perturbed boundary layer humidity, based on Tompkins and Berner 2008, *DOI: 10.1029/2007JD009284)*
- All use stochastic random pattern generators to generate spatially and temporally correlated noise.

Examples of stochastic patterns



(from M Leutbecher)

ECMWF method (SPPT)

Stochastically Perturbed Physics Tendency

- Perturbed Physics tendencies

$$X_p = (1 + r\mu) X_c$$

Original physics tendencies

μ: vertical weight: 1.0 between surface and 100 hPa, decays to zero between 100 hPa and 50 hPa.

r: horizontal weights: range from -1.0 to 1.0, a red noise process with a

- temporal timescale of 6 hours
- e-folding spatial scale of 500 km

Stochastic boundary-layer humidity

- SPPT only modulates existing physics tendency (cannot change sign or structure, trigger new convection).
- Triggers in convection schemes very sensitive to BL humidity.

$$q_{perturbed} = (1 + r\mu)q$$

- Vertical weight **r** decays exponentially from surface. Added every time step after physics applied. Random pattern **μ** has a (very small) amplitude of 0.0015, horizontal/temporal scales same as SPPT.

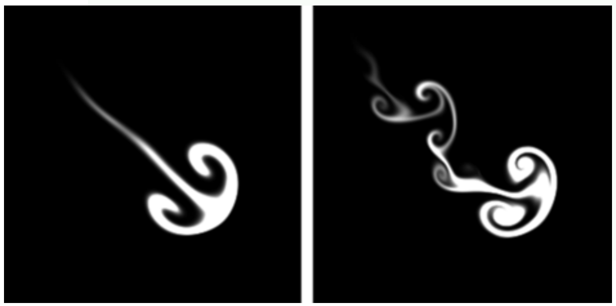
Stochastic Kinetic Energy Backscatter

- Algorithm described in Shutts (2005), Berner et al (2009)
  - Designed to represent the effects of dissipated motions near truncation scale on resolved motions.
  - Random patterns are modulated by amplitude of KE dissipation (numerical, possibly other sources like convection – we only consider numerical dissipation here).

Vorticity confinement

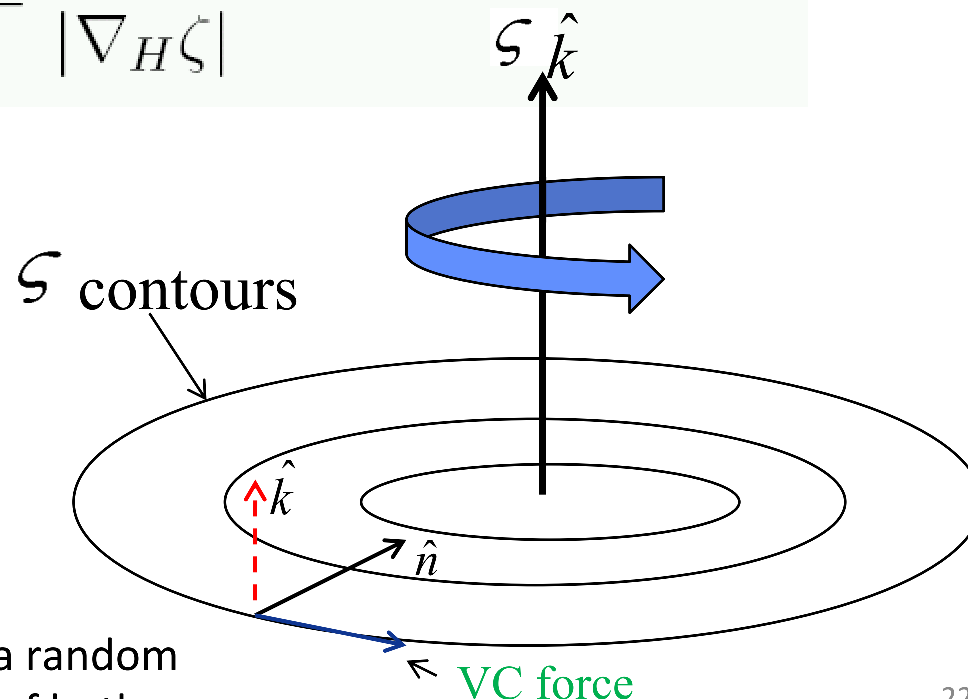
(Sanches, Williams and Shutts, 2012 QJR doi 10.1002)

$$\frac{D\mathbf{V_H}}{Dt} + f\mathbf{k} \times \mathbf{V}_H + \nabla\phi = \mu\nabla^2\mathbf{V}_H + \epsilon\hat{\mathbf{n}} \times |\zeta|\,\hat{\mathbf{k}}$$



$$\hat{\mathbf{n}} = \frac{\nabla_H \zeta}{|\nabla_H \zeta|}$$

Figure 6: Two frames of animation from two mpeg movies created using `flowanim` and `mpeg2encode`. Both frames depict the 60th frame of the movie. The left animation is created without vorticity confinement, the one on the right with vorticity confinement and a *relatively* high force factor

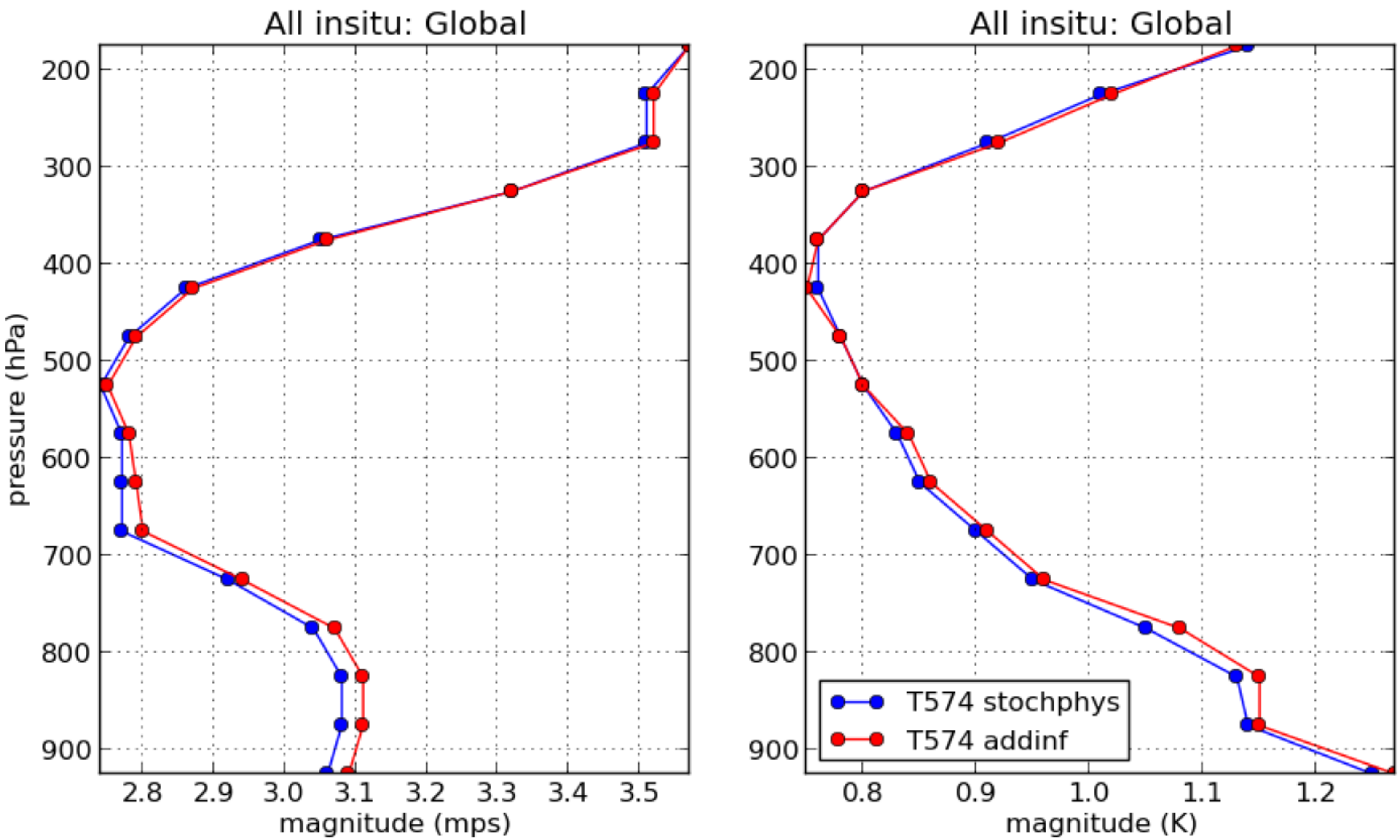


$\epsilon\hat{\mathbf{n}}$ acts as an advective velocity

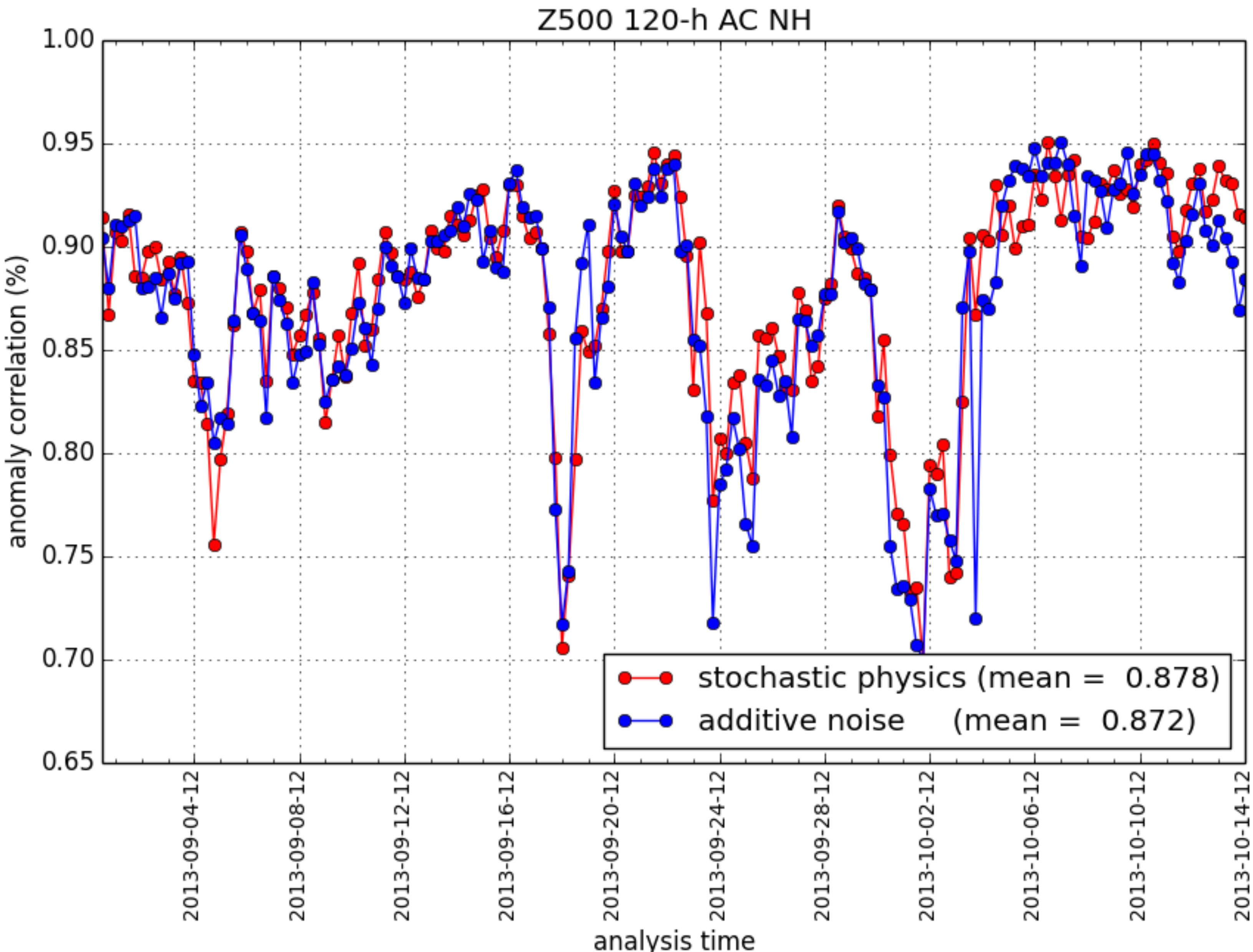
ε can be a constant (determinstic VC) or a random pattern (stochastic VC) or a combination of both.

Additive noise vs stochastic physics

Vector Wind (left) and Temp (right) O-F (2013091000-2013101412)



Additive noise vs stochastic physics

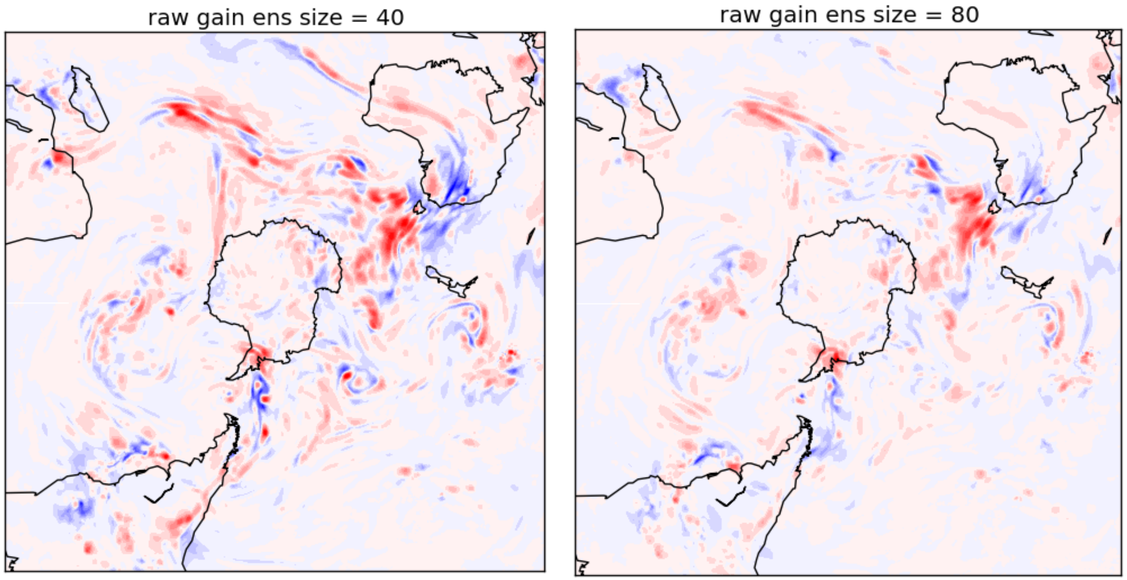


Factors limiting EnKF performance

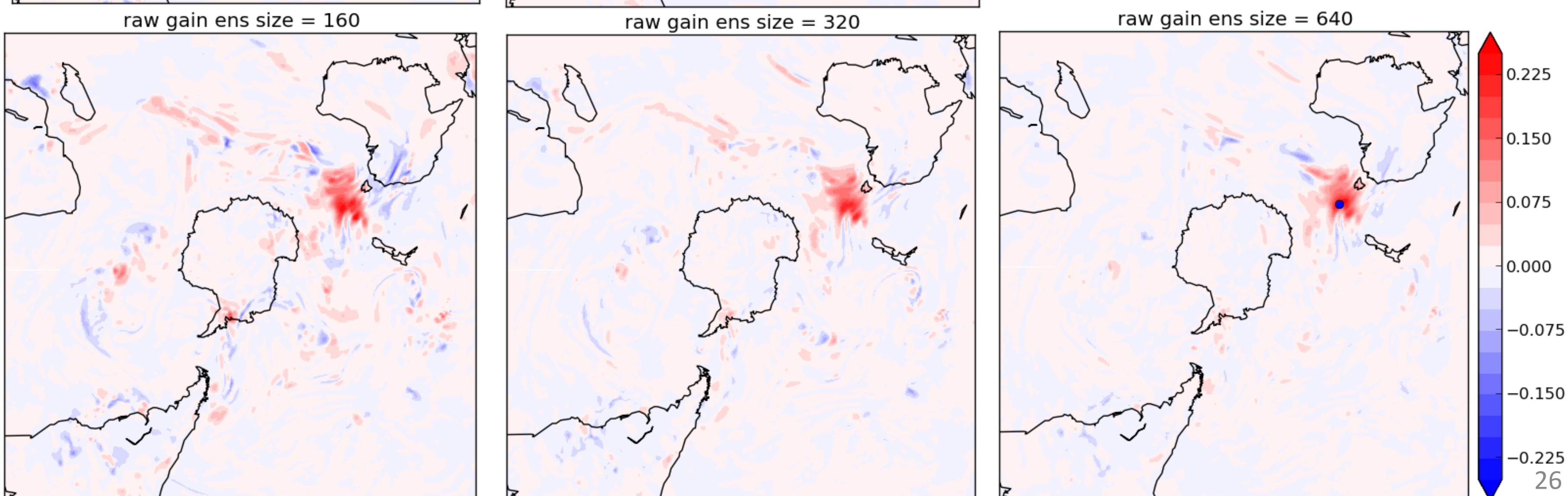
2) Treatment of sampling error (localization)

- All EnKF implementations localize the spatial impact of observations on the model state.
- Done by spatially modulating covariance between obs. prior and model state, or by only using observations ‘close’ to a model state variable to update that variable.
- Needed to account for low rank of ensemble (compared to model state).
- Methods used currently are not flow dependent, and assume there is no sampling error at ob location.

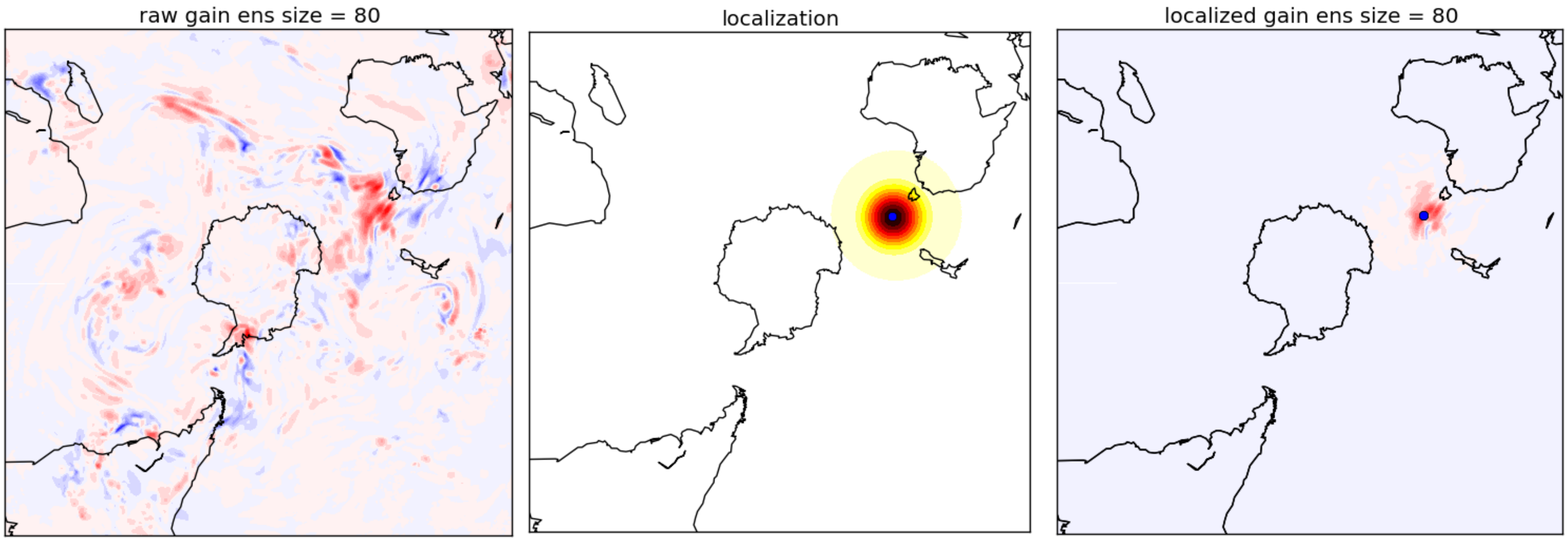
Covariance localization



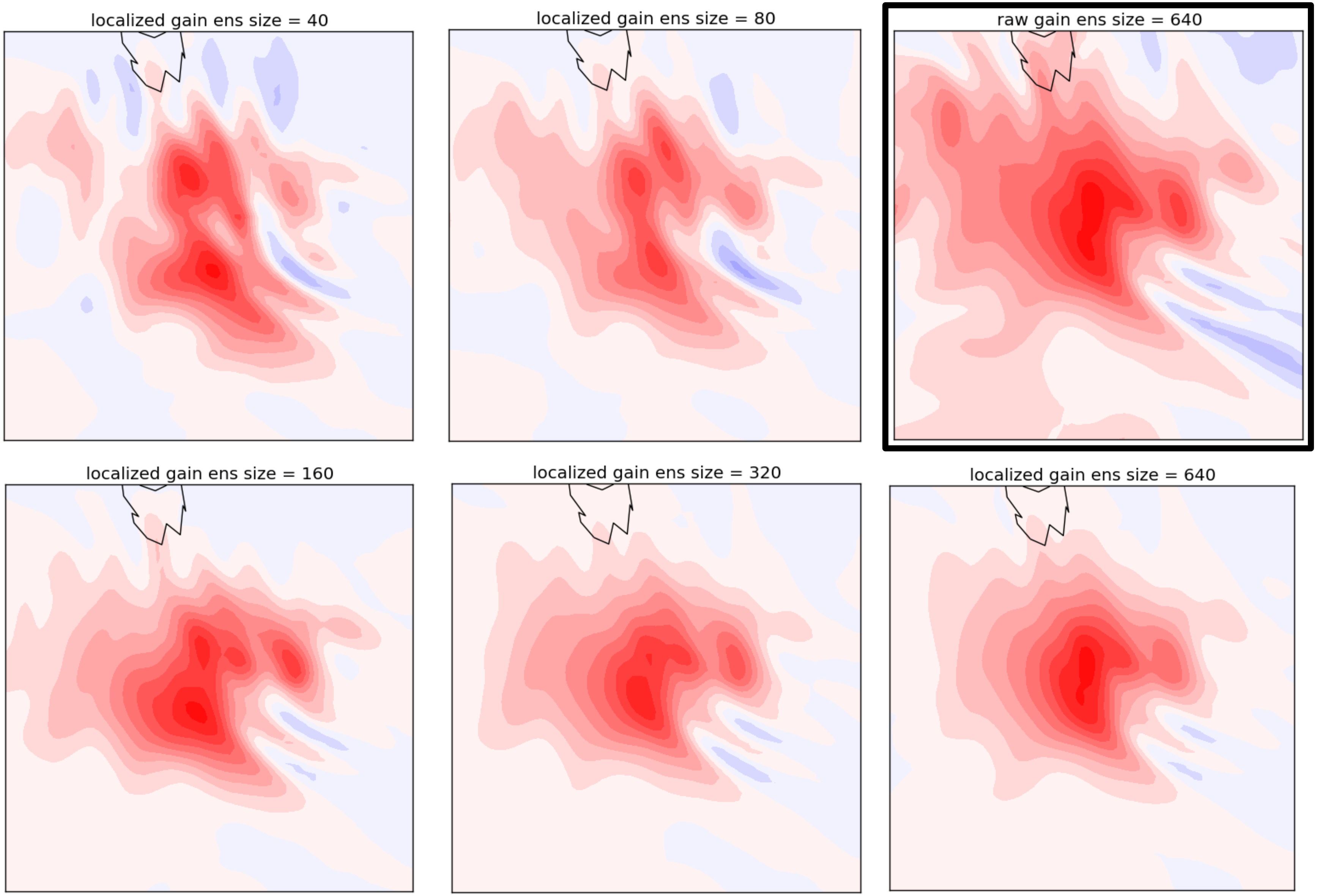
- AMSUA_n15 channel 6 radiance at 150E,-50S.
- Increment to level 30 (~310mb) temperature for a 1K O-F for 40,80,160,320 and 640 ens members with no localization.
- Ens generated by running with stochastic physics from operational analysis.



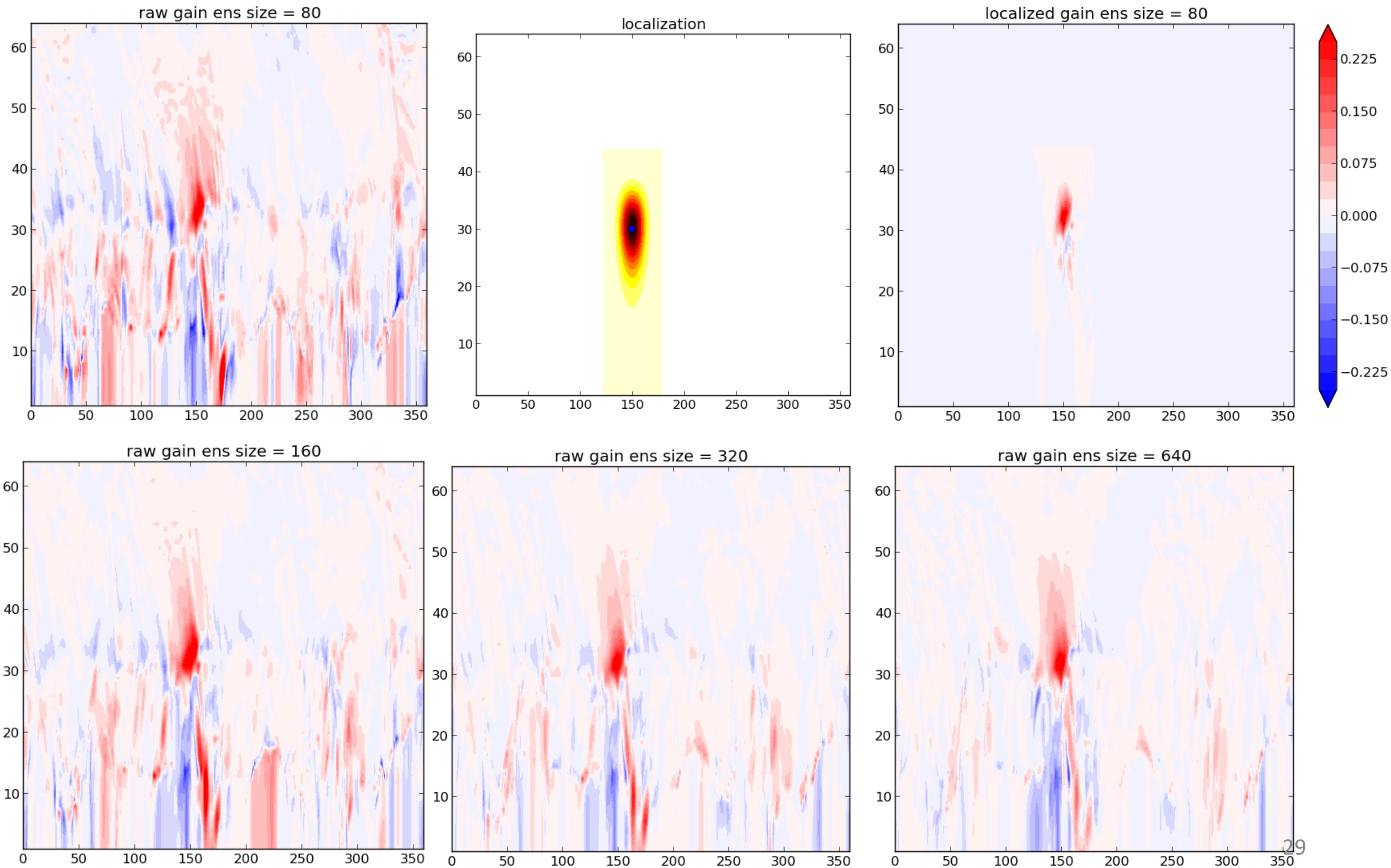
Covariance Localization



Localized covariances



Vertical localization



Ensemble-Var methods: nomenclature

- ***En-4DVar***: Use EnKF to propagate $\mathbf{P^b}$ from one assimilation window to the next (replace static $\mathbf{P^b}$ with EnKF estimate of $\mathbf{P^b}$ at start of 4DVar window).
- ***4D-EnVar***: $\mathbf{P^b}$ at every time in the assim. window comes from ensemble estimate (linear model no longer used).
- As above, with ***hybrid*** in name: $\mathbf{P^b}$ is a linear combination of static and ensemble components.
- ***3D-EnVar***: same as 4D ensemble Var, but $\mathbf{P^b}$ is assumed to be constant through the assim. window (current NCEP implementation).

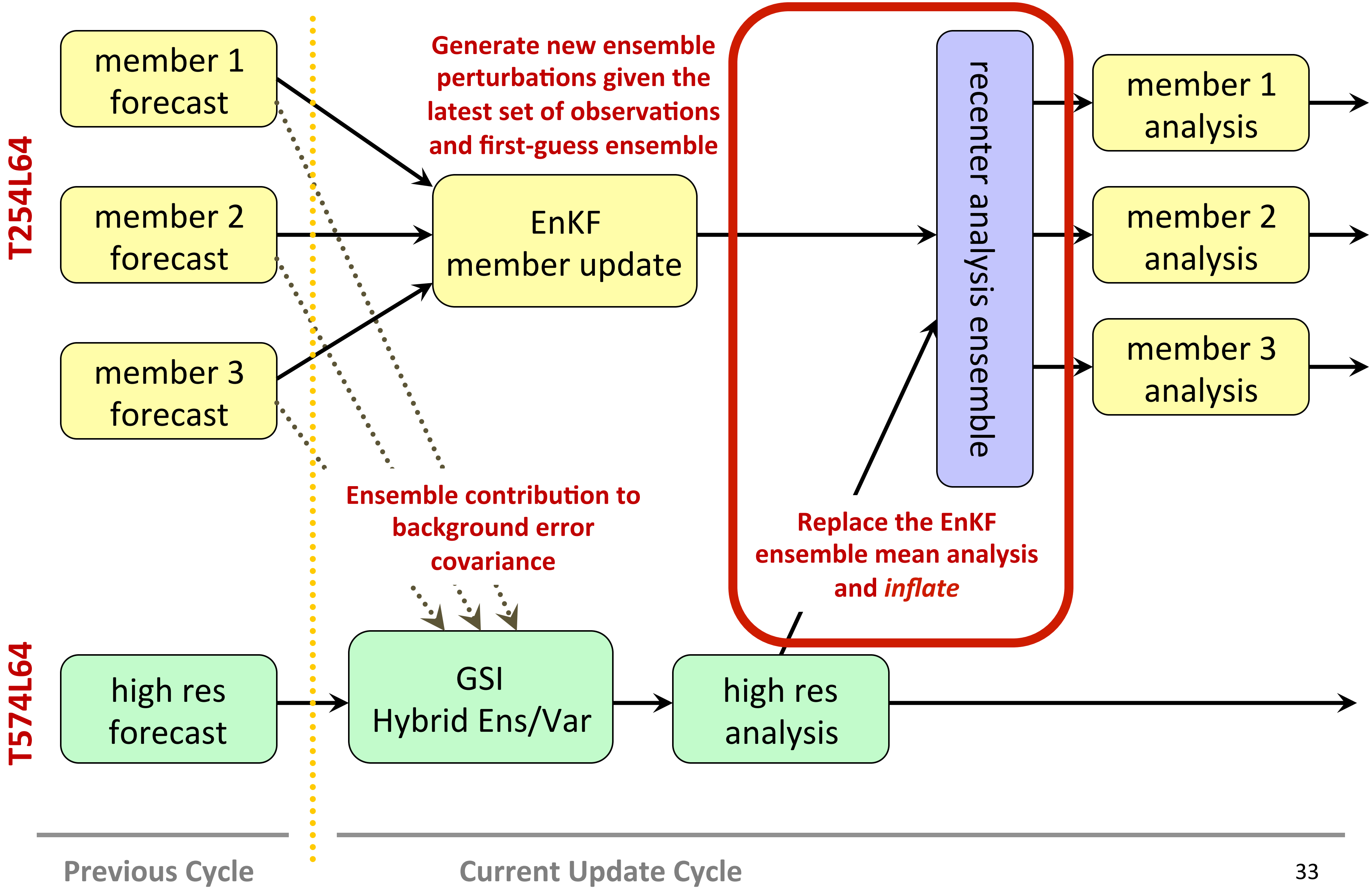
Why combine EnKF and Var?

Features from EnKF	Features from Var
Can propagate $\mathbf{P}^b$ from across assimilation windows	Treatment of sampling error in ensemble $\mathbf{P}^b$ estimate does not depend on **H**.
More flexible treatment of model error (can be treated in ensemble)	Dual-resolution capability – can produce a high-res "control" analysis.
Automatic initialization of ensemble forecasts.	Ease of adding extra constraints to cost function, including a static $\mathbf{P}^b$ component.
$\mathbf{P}^b$ estimated with a much better model (full fcst model)	No sampling error in evolution of $\mathbf{P}^b$ within assim. window (when TLM is used)

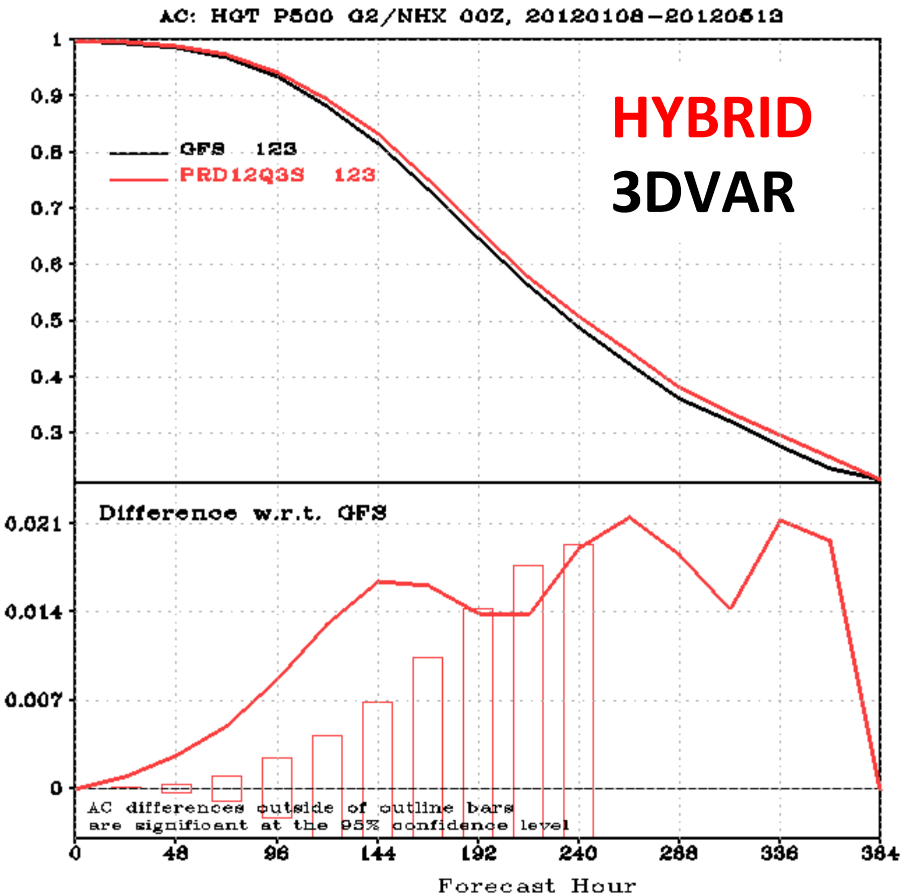
Current implementations

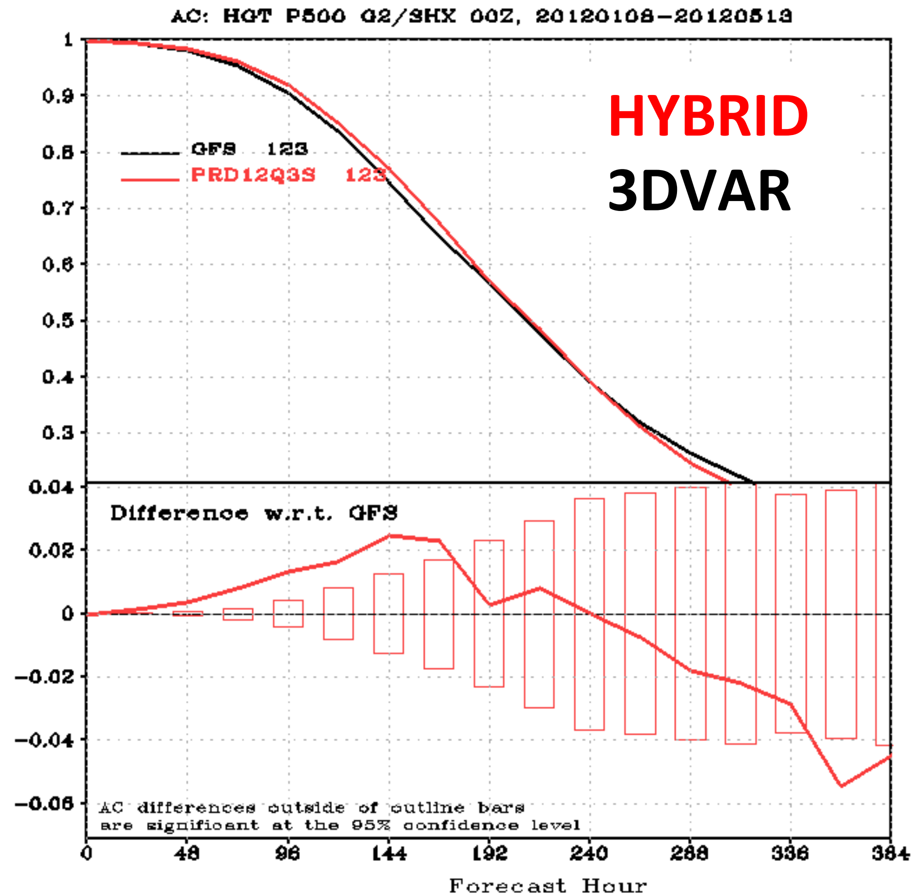
- ***NCEP*** uses hybrid 3D-EnVar (since May 2012), with 80 member ESRL/PSD serial EnKF implementation. ¼ static, ¾ ensemble $\mathbf{P}^b$.
- ***UKMet*** uses hybrid En-4DVar (DOI: 10.1002/qj.2054) since late 2011, with 23 member Ensemble-transform Kalman Filter (ETKF). ½ static and ½ ensemble in $\mathbf{P}^b$. Will be switching to 4D-EnVar.
- ***Env Canada*** uses hybrid 4D-EnVar (since mid 2013), with 192 member EnKF with ½ static and ½ ensemble.
- ***ECMWF*** uses ensemble of low-res 4DVar analyes to define diagonal elements of $\mathbf{P}^b$ in higher-res 4DVar with 12-h window. Moving to very long window (2-3 days) to improve estimate of $\mathbf{P}^b$.

NCEP hybrid 3D-EnVar



500mb Anomaly Correlation pre-implementation real time

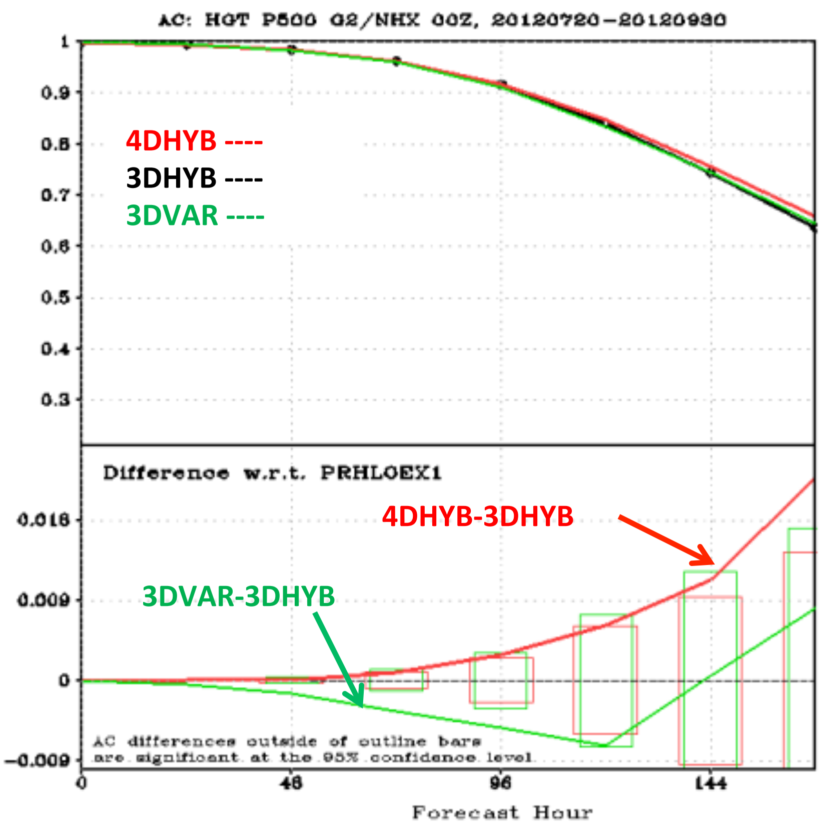




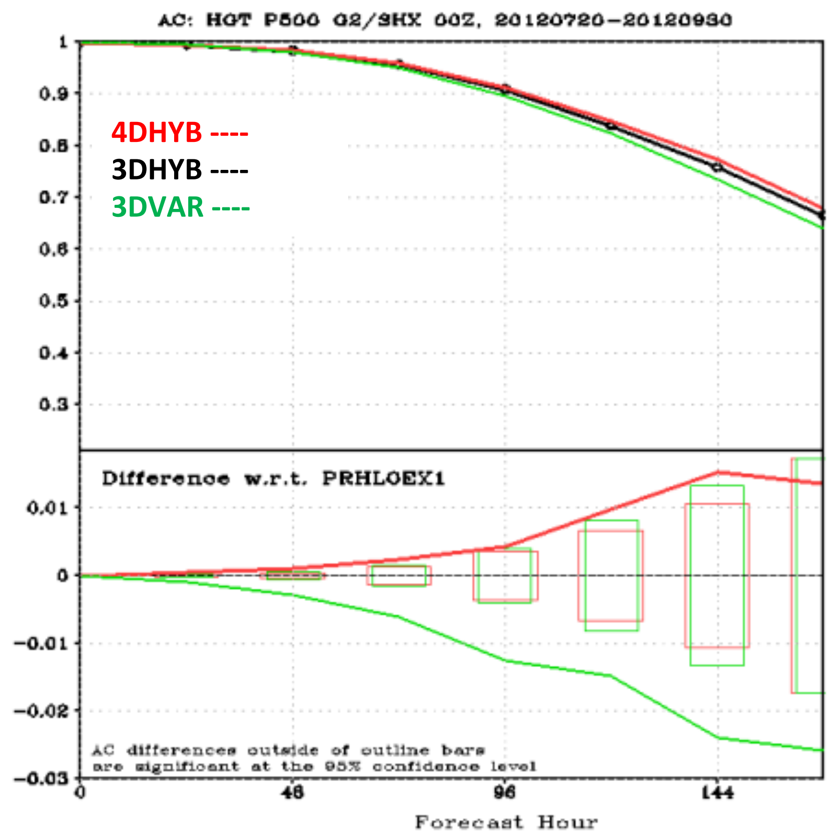
4D-EnVar vs 3D-EnVar (and 3DVar)

courtesy D. Kleist

Northern Hemisphere



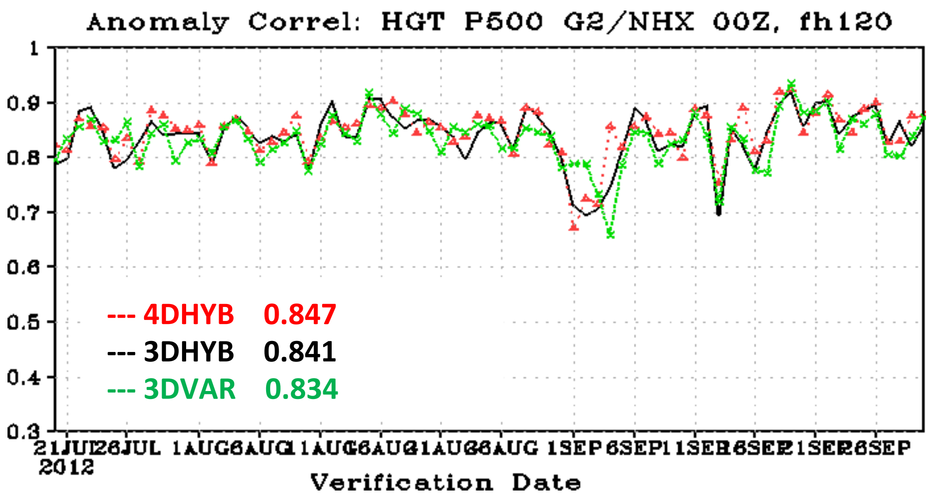
Southern Hemisphere



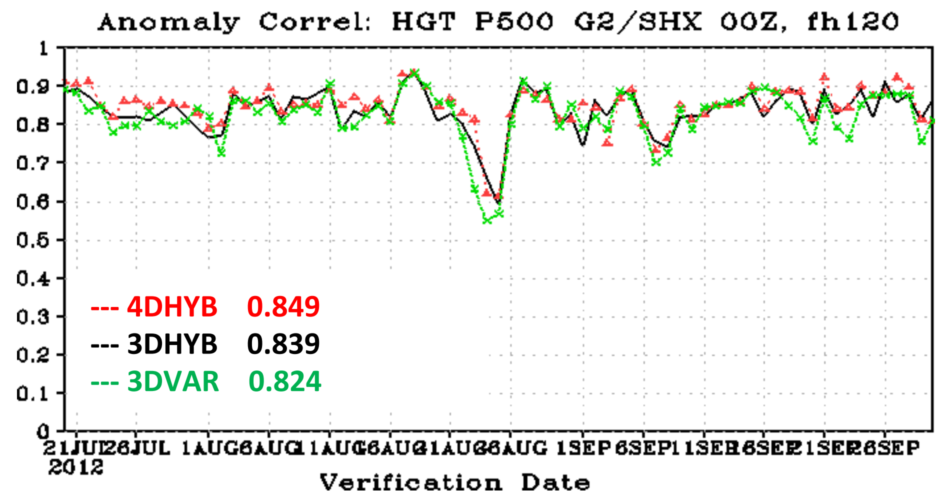
Move from 3D Hybrid (current operations) to Hybrid 4D-EnVar yields improvement that is about 75% in amplitude in comparison from going to 3D Hybrid from 3DVAR.

500 hPa Day 5 Time Series

Northern Hemisphere

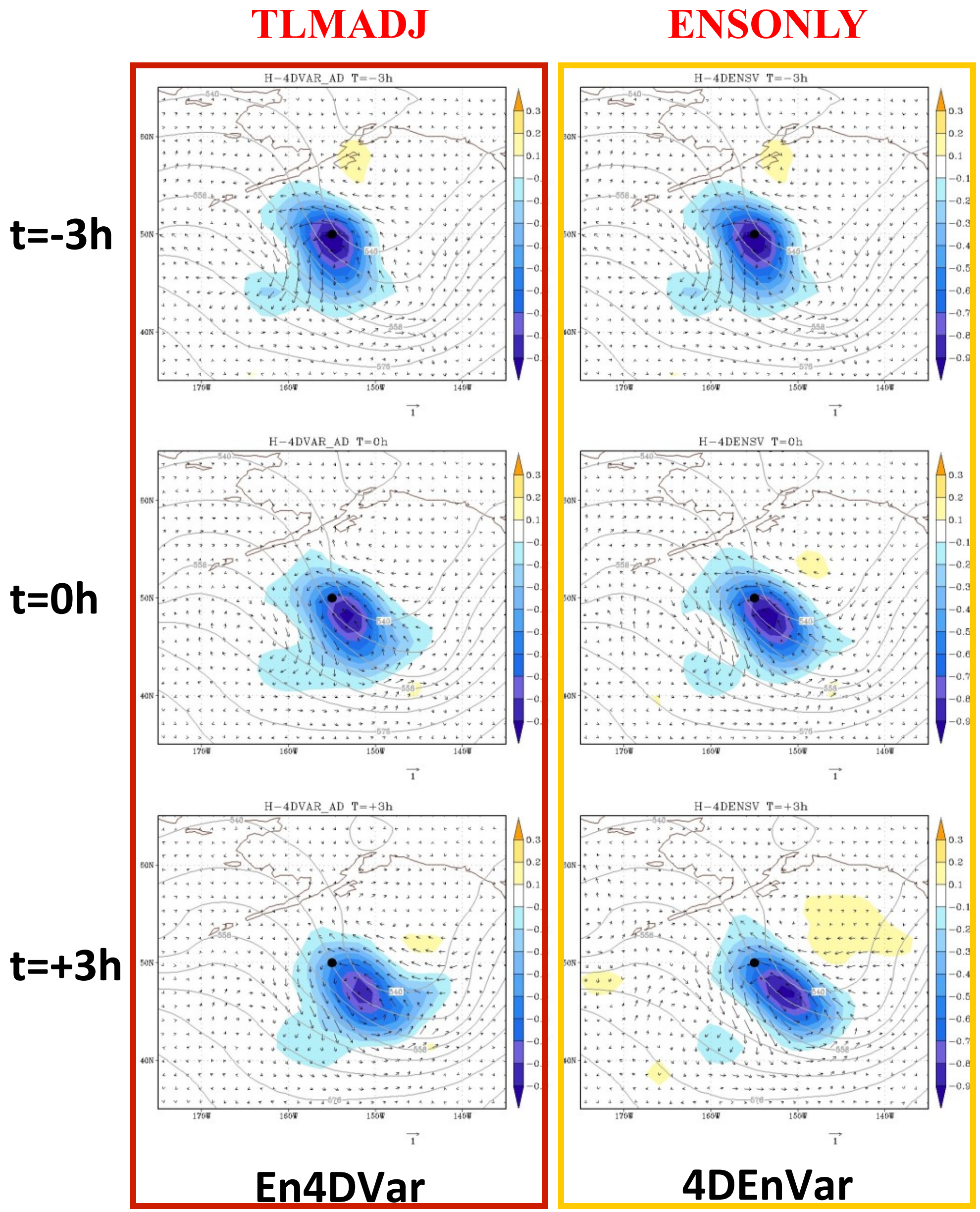


Southern Hemisphere



Again, going from 3D to 4D-hybrid yields improvement at day 5 similar (not quite as large in SH) to what was seen in going from 3DVAR to 3D-hybrid

Time Evolution of Increment: hybrid 4D-EnVar vs En-4DVar

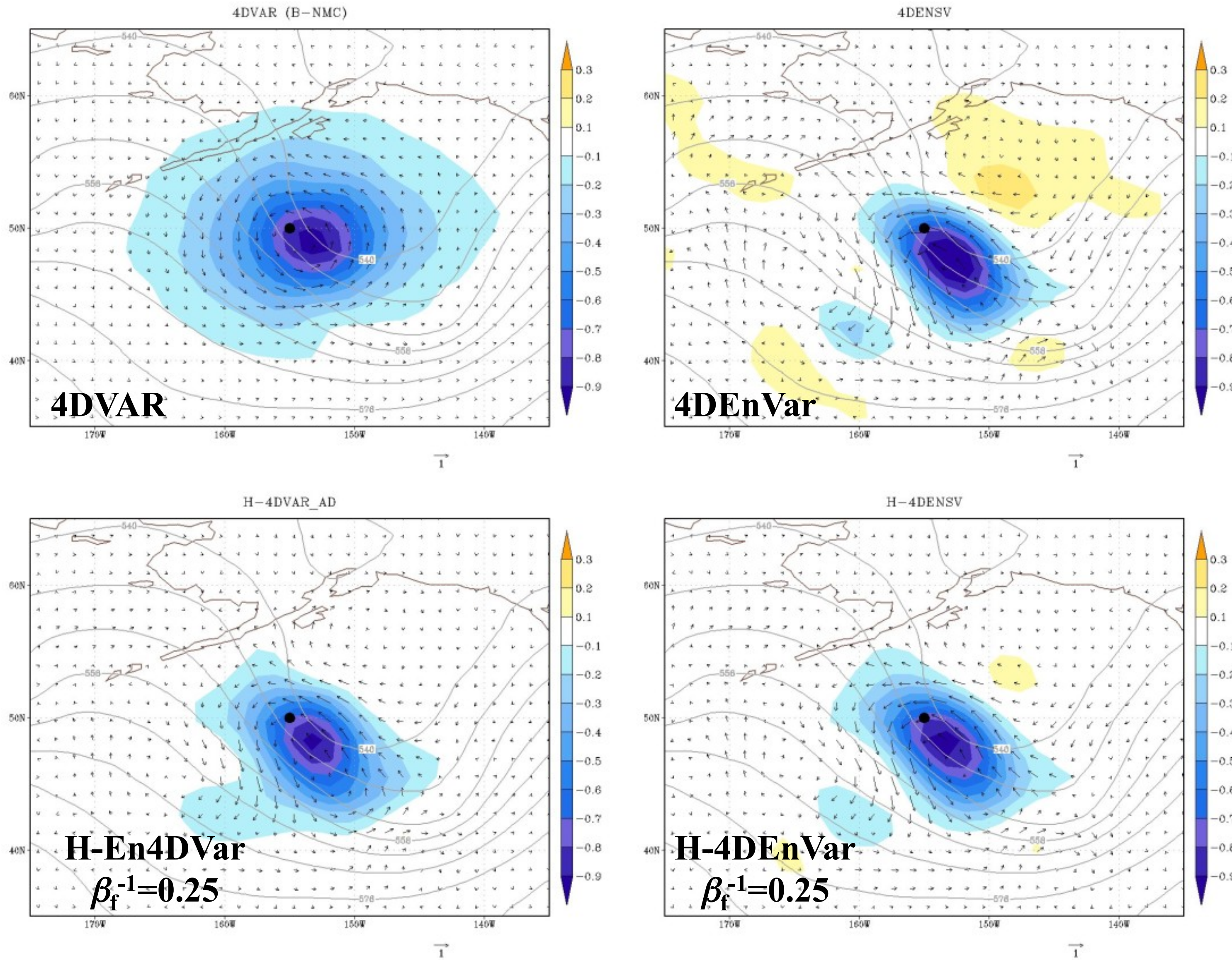


Solution at beginning of window same to within round-off (because observation is taken at that time, and same weighting parameters used)

Evolution of increment qualitatively similar between dynamic and ensemble specification

** Current linear and adjoint models in GSI are computationally impractical for use in 4DVAR other than simple single observation testing at low resolution

Single Observation (-3h) Example for 4D Variants



Summary

- Improvements in analysis quality in NWP over the last 20 years have resulted mainly from:
  - Improvements to forecast model.
  - Improving $\mathbf{P}^b$ estimates using 4D-Var and ensembles.
- Further improvements to come via:
  - Larger ensembles.
  - Better treatment of model and sampling error.
  - Methods to deal with non-Gaussian error statistics (from e.g. feature displacement, monotonicity).

Extra slides

What if ob errors are correlated?
(**R** not diagonal)

Either thin obs so that they are space far enough apart so that errors are uncorrelated, or….

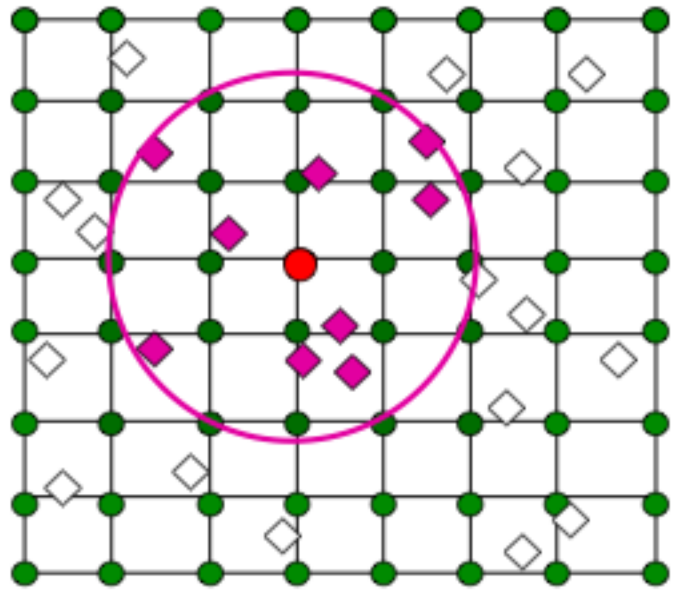
- Diagonalize $\mathbf{R} = \mathbf{S}^{T}\mathbf{\Omega}\mathbf{S}$, where $\mathbf{\Omega}$ is diagonal.
- Replace $\mathbf{H}$ with $\mathbf{G}=\mathbf{HS}$.
- Replace $\mathbf{Y}^{b}$ with $\mathbf{Z}^{b}=\mathbf{SY}$, $\mathbf{y}^{o}$ with $\mathbf{z}^{o}=\mathbf{Sy}^{o}$, $\mathbf{R}$ with $\mathbf{\Omega}$.
- Assimilate obs serially in this transformed space.

LETKF Algorithm

Perform data assimilation in a local volume, choosing observations

The state estimate is updated at the central grid red dot

All observations (purple diamonds) within the local region are assimilated



Ob error in local volume is increased as a function of distance from red dot, reaching infinity at edge of circle.

Local Ensemble Transform Kalman Filter (LETKF)

Globally:

Forecast step: $\mathbf{x}^b_{n,k} = M_n\left(\mathbf{x}^a_{n-1,k}\right)$

Analysis step: construct $\mathbf{X}^b = \left[\mathbf{x}^b_1 - \overline{\mathbf{x}}^b \mid \ldots \mid \mathbf{x}^b_K - \overline{\mathbf{x}}^b\right]$;

$$\mathbf{y}^b_i = H(\mathbf{x}^b_i);\ \mathbf{Y}^b_n = \left[\mathbf{y}^b_1 - \overline{\mathbf{y}}^b \mid \ldots \mid \mathbf{y}^b_K - \overline{\mathbf{y}}^b\right]$$

Locally: Choose for each grid point the observations to be used, and compute the local analysis error covariance and perturbations in ensemble space:

$$\tilde{\mathbf{P}}^a = \left[(K-1)\mathbf{I} + \mathbf{Y}^{bT}\mathbf{R}^{-1}\mathbf{Y}^b\right]^{-1};\ \mathbf{W}^a = [(K-1)\tilde{\mathbf{P}}^a]^{1/2}$$

Analysis mean in ensemble space: $\overline{\mathbf{w}}^a = \tilde{\mathbf{P}}^a\mathbf{Y}^{bT}\mathbf{R}^{-1}(\mathbf{y}^o - \overline{\mathbf{y}}^b)$

and add to $\mathbf{W}^a$ to get the analysis ensemble in ensemble space

The new ensemble analyses in model space are the columns of $\mathbf{X}^a_n = \mathbf{X}^b_n\mathbf{W}^a + \overline{\mathbf{x}}^b$. Gathering the grid point analyses forms the new global analyses. Note that the the output of the LETKF are analysis weights $\overline{\mathbf{w}}^a$ and perturbation analysis matrices of weights $\mathbf{W}^a$. These weights multiply the ensemble forecasts.

Relaxation To Prior Spread (RTPS) Inflation

Described in DOI: 10.1175/MWR-D-11-00276.1

Inflate posterior spread (std. dev) σ^a back toward prior spread σ^b

$$\sigma^a \leftarrow (1-\alpha)\sigma^a + \alpha\sigma^b$$

Equivalent to

$$\mathbf{x}_i^{'a} \leftarrow \mathbf{x}_i^{'a}\sqrt{\alpha\frac{\sigma^b-\sigma^a}{\sigma^a}+1}$$